



المادة: Data Analysis in R

المدة: ساعتين

الأستاذ: د. حسين هزيمة

المرحلة: الإجازة

السنة المنهجية: الثانية

الاختصاص: علم البيانات - Data Science

Part 1: Multiple Choice Questions (17)

1. Which of the following best describes time series analysis?
A. Clustering data into groups
 B. Analyzing data collected over time intervals
C. Compressing data files
D. Encrypting datasets
2. Which component represents a consistent increase or decrease in time-series data?
A. Noise
 B. Trend
C. Frequency
D. Variance
3. A moving average is mainly used to:
A. Increase randomness
 B. Smooth out fluctuations in data
C. Encrypt time-series values
D. Classify text documents
4. Which moving average gives more importance to recent observations?
 A. SMA
B. EMA
C. WMA
D. CMA
5. In linear regression, the dependent variable is usually represented by:
 A. X
B. Y
C. Z
D. T
6. The regression equation is commonly written as:
 A. $y = mx + b$
B. $x = my + b$
C. $y = m/b$
D. $y = x^2 + b$
7. In cosine similarity, a similarity value close to 1 means:
A. Documents are very different

- B. Documents are unrelated
 C. Documents are highly similar
D. Documents are empty
8. Which similarity technique is based on the angle between vectors?
A. Jaccard similarity
B. Euclidean distance
 C. Cosine similarity
D. Manhattan distance
9. TF-IDF is mainly used in:
A. Image compression
 B. Text mining
C. Network routing
D. Video streaming
10. Which of the following is a stop word?
A. Computer
B. Science
 C. The
D. Database
11. Jaccard similarity is calculated using:
A. Dot product only
 B. Intersection divided by union
C. Matrix multiplication
D. Average frequency
12. Edit distance measures:
A. The size of a database
 B. Similarity between strings
C. Network centrality
D. Forecast accuracy
13. Which operation is NOT part of edit distance computation?
A. Insertion
B. Deletion
C. Substitution
 D. Multiplication
14. Which data type has no true zero?
A. Ratio
B. Nominal
 C. Interval
D. Ordinal
15. Which of the following is an example of ratio data?
A. Temperature in Celsius
B. Eye color
C. Ranking
 D. Weight
16. Exploratory Data Analysis (EDA) is mainly used to:
A. Confirm hypotheses only
 B. Discover patterns and insights

- C. Encrypt information
 - D. Remove databases
17. Which forecasting model considers trend and seasonality together?
- A. KNN
 - B. Winters Model
 - C. Jaccard Model
 - D. BFS Model

Part 2: True / False Questions (17)

1. Time-series data is collected over intervals of time. T
2. A stationary series contains a clear increasing trend. F
3. EMA gives equal weights to all observations. F ?
4. Linear regression predicts the value of Y from X. T
5. More data can improve regression results. T
6. Cosine similarity depends on the angle between vectors. T
7. Two identical vectors have cosine similarity equal to 1. T
8. TF-IDF is used in textual data analysis. T
9. Jaccard similarity uses the union and intersection of sets. T
10. Edit distance can include insertion and deletion operations. T
11. Nominal data can be ordered numerically. F
12. Ratio data contains a true zero value. T
13. CDA stands for Confirmatory Data Analysis. T
14. Stop words are usually removed in text preprocessing.
15. Weighted Moving Average assigns different weights to observations. F
16. Cosine similarity can be negative for identical vectors. F
17. Trend and seasonality are components of time-series analysis. T

Part 3: Problem Solving Questions (2)

Problem 1: Time-Series and Regression Analysis

Given the following quarterly sales data for a company:

Quarter	Sales (Y)	Advertisement Budget (X)
Q1	50	10
Q2	60	15
Q3	65	18
Q4	80	25

1. Calculate the Simple Moving Average (SMA) using a window size of 2.
2. Find the linear regression equation $y = mx + b$.
3. Predict the sales value when the advertisement budget is 30.

Problem 2: Text Similarity and Vector Space Model

Given the following document collection:

- D1 = data science machine learning
 - D2 = machine learning artificial intelligence
 - D3 = data mining science learning
 - D4 = artificial intelligence deep learning
1. Construct the binary term-document matrix.
 2. Construct the term-frequency matrix.
 3. Calculate the cosine similarity between D1 and D2.
 4. Calculate the cosine similarity between D1 and D3.
 5. Determine which document is most similar to D1.
 6. Draw the vectors in a 2D vector space and illustrate the angle relationship between them.